PCR Template Preparation for Capillary DNA Sequencing

BioTechniques 30:537-542 (March 2001)

ABSTRACT

Fluorescence-based capillary DNA sequencing has facilitated the early completion of several complex sequencing projects. While capillary systems offer great benefits in terms of ease of use and automation, we find that they are sufficiently different from slab gel separation methodologies, demanding re-examination of the protocols used to generate and use DNA sequencing templates. We have recently initiated a large-scale Human Open Reading Frame™ EST project involving 30 laboratories feeding 11 MegaBace™ 1000 capillary sequencers. The group has already produced more than 300000 valid sequences. The most successful template preparation protocol we have found is described here. We have found that a crucial step is the standardization of the quantity and quality of the templates, which have been achieved by overnight bacterial culture followed by PCR using limiting amounts of primers. Using this protocol, there is no need for post-PCR purification, and the final preparation cost is US $0.09/template. After sequencing 10 848 templates using this protocol, 78% of the reads were accepted (after discarding vectors without inserts and inserts smaller than 100 nucleotides), and 85% of the total number of bases had Phred scores of 15 or above.

INTRODUCTION

Improvements in DNA sequencing technology have dramatically increased the pace of data generation. In particular, capillary DNA sequencers have permitted the automation of a number of steps, reducing electrophoresis time and eliminating slab gel casting, sample loading, and lane tracking. These advances resulted in the completion of the genome-sequencing phase of *Drosophila melanogaster* in just four months (1). The same technology is being used for the rapid completion of the human genome sequence.

We have recently initiated a sequencing project, the FAPESP/LICR Human Cancer Genome Project (HCGP), with the aim of generating 1 million expressed sequence tags (ESTs) derived from human tumors (2). The ESTs obtained are derived from the central coding regions of human expressed genes by using Open Reading frame ESTs (ORESTES), a novel approach to cDNA sequencing (3). The HCGP is one of the activities of the Brazilian ONSA Virtual Genomics Institute (12). The HCGP network is composed of 30 sequencing laboratories feeding 11 MegaBace™ 1000s (Amersham Pharmacia Biotech, Piscataway, NJ, USA). For the success of our endeavor, it was necessary to develop a method for template preparation that resulted in high-quality sequences, that was cost effective (in terms of time and reagents required for template preparation), and that permitted a high percentage of successful reads. The initial phase of the project was dedicated to comparing template preparation strategies, operator training issues, and sequencing protocol development. The most successful approach to template preparation is described here.

MATERIALS AND METHODS

Template Preparation and Sequencing

Individual bacterial colonies (JM109 or DH5α), transformed with pUC18 containing human cDNA inserts, were picked from agar plates using sterile toothpicks and transferred to flat-bottom 96-well plates containing 120 µL 2× YT or LB media supplemented with 100 µg/mL ampicillin. Plates were wrapped with vitafilm, and the clones were grown at 37°C, without shaking, for 20–24 h. Samples (5–10 µL) of the resulting bacterial culture were diluted 10 times in water, and 1 µL was used to generate the sequencing templates by PCR. Individual amplification reactions contained 1× PCR buffer [25 mM Tris-HCl, pH 8.3, 75 mM KCl, and 10 mM dithiothreitol (DTT)], 125 µM dNTPs, 1.5 mM $MgCl_2$, 1 pmol each M13 universal primer, and 1 U *Taq* DNA polymerase (Life Technologies, Rockville, MD, USA) in a final volume of 15 µL. A large volume of PCR mixture could be prepared for many hundred reactions, and DNA polymerase could be added just before the reaction setup. The cycling conditions started with a longer denaturation step at 95°C for 4 min, followed by 35 cycles of 95°C for 45 s, 55°C for 45 s, and 72°C for 1 min. A final extension step at 72°C for 5 min was undertaken. A random inspection of a few templates from each amplification batch was undertaken by visualizing 3 µL of the PCR products on 1% agarose gels stained with ethidium bromide.

The cycle sequencing of PCR-generated templates was performed with the traditional dideoxy chain termination chemistry (11) of PCR-generated templates using 8 µL DYEnamic™ E T Dye Terminator cycle sequencing kit (Amersham Pharmacia Biotech), 2.5 pmol of one of the same primers used for template generation, and 1 µL of the PCR products. The termination mixture of the dye terminator kit contains a thermostable DNA polymerase, enzyme buffer, each of the four dideoxy terminators labeled with fluorescein, and one of four different rhodamine dyes (9,10). Water was used to complete the final sequencing reaction volume to 20 µL. Cycling conditions were 95°C for 2 min, followed by 35 cycles of 95ºC for 10 s, 50°C for 15 s, and 60°C for 1 min. The reaction was precipitated by adding 2 µL ammonium acetate and 50 µL ice-cold absolute ethanol. The components were vortex mixed, and the reaction was allowed to stand for 15 min at room temperature and centrifuged for 30 min at 3100× *g*. The supernatant was removed, and the plate was briefly centrifuged upside down over a paper towel for a few seconds. The pellet was resuspended in 20 µL formamide loading buffer (supplied with the ET sequencing kit), vortex mixed, and briefly centrifuged. All the reactions are undertaken in 96-well plates for the simultaneous sequencing in the capillary DNA sequencer MegaBace 1000. The sequencing components were electro-injected for 90 s at 1 kV. Sample denaturation before loading was not required.

The effects of a post-PCR clean up before the sequencing was evaluated. For this, the same 192 templates were sequenced with or without treatment

with exonuclease I and shrimp alkaline phosphatase (*Exo*I/SAP) after PCR. This treatment eliminates primers and dNTPs, previously used for template generation, from the sequencing reaction. The *Exo*I/SAP treatment was conducted by mixing the cycle sequencing products with 10 U *Exo*I and 2 U SAP, included in the PCR product pre-sequencing kit from USB (Cleveland, OH, USA), and incubating at 37°C for 15 min. After treatment, the enzymes were denatured for 15 min at 80°C, and the treated DNA was used directly for sequencing as described above.

Sequence Analysis

The chromatograms generated after sequencing the PCR templates were submitted to an automated protocol for base calling, sequence quality assignment, and trimming vector and primer sequences. The chromatograms derived from the capillary DNA sequencers were exported in the Appied Biosystems Data (ABD) format and analyzed using the Phred software (5,6). Error probabilities are assigned from the chromatograms by Phred analysis, based on parameters such as background noise, uncalled/called base ratios, peak spacing, and resolution. In our project, a sequence is accepted if it contains, after removing vector and primer sequences, at least 100 nucleotides with Phred quality of 15 or above (this score indicates that the base call has an error probability of 3.16%).

RESULTS AND DISCUSSION

The results of sequencing 10 848 templates from 113 96-well plates, using the protocol described, were evaluated in terms of the number of accepted reads and the number of bases with Phred scores of 15 or above. Seventy-eight percent of the sequences generated were accepted using the quality parameters described in the previous section. From the total number of bases sequenced, 84.9% had Phred values above 15. During the complete evaluation period, the percentage of accepted reads was greater than 70%, reaching peaks of 90%. The percentage of bases with Phred values above 15 was around 80% throughout the period, also reaching 90% in some cases (Figure 1).

As the sequencing templates described here were derived from PCR products, the size of the PCR fragments is an absolute limiting factor in the final length of the sequence obtained. In this study, this factor was found to be the principal limiting factor in read length. The average size of the PCR templates in this series of experiments was 452 nucleotides, and the average read length/template (Phred scores above 15) was 404 nucleotides. When plates containing inserts of 270 nucleotides were sequenced, around 250 nucleotides of high-quality sequences were obtained (plates 46–48). In contrast, when inserts of about 630 nucleotides were sequenced, high-quality reads of about 600 nucleotides were obtained (plates 76–78). Figure 2 shows the relationship between the template size and the average number of high-quality reads. The overall percentage of high-quality reads is high and consistent.

To evaluate the effect of post-PCR cleanup, which is usually performed to eliminate unincorporated oligonucleotide primers and dNTPs, replicates of two plates (192 samples) were sequenced with or without previous *Exo*I/SAP treatment (PCR product pre-sequencing kit). Treatment was performed in 5 µL PCR products, using 10 U *Exo*I and 2 U SAP for 15 min at 37°C. Enzymes were subsequently denatured at 80°C for 15 min. DNA sequencing was performed and evaluated as described above. The average insert size of these clones was 517 nucleotides, and the average read length with or without treatment was 487 and 445 nucleotides, respectively. We conclude that for our purposes the *Exo*I/SAP treatment is not cost effective.

Salt and template concentrations in the sequencing reaction products are critical to the success of capillary DNA sequencing. The concentration of salt present in different PCR products is constant. Therefore, if the template con-

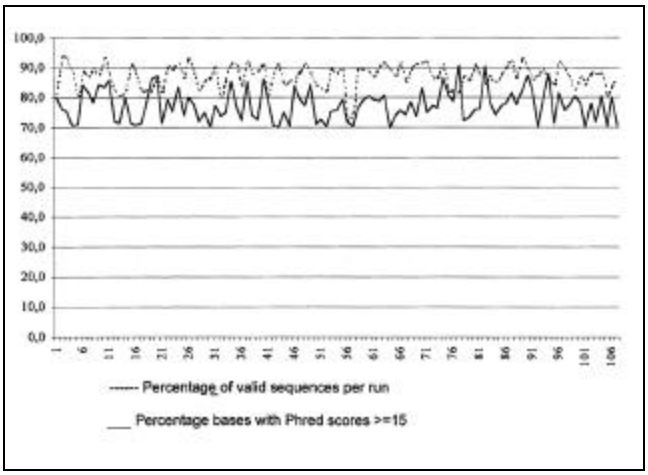

Figure 1. Evaluation of sequencing performance in terms of the number of valid reads and the number of high-quality bases. Analysis of 113 sequencing plates is presented in terms of the average percentage of valid sequences per plate (dashed line) and percentage bases with Phred scores greater than or equal to 15 (solid line).

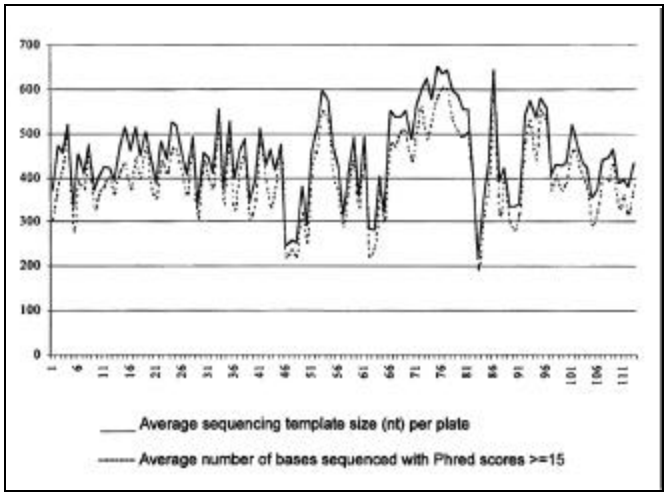

Figure 2. Correlation between the size of sequencing templates and the number of high-quality bases. The average size of sequencing templates is shown as the solid line and number of bases with Phred scores greater than or equal to 15 are shown as the dashed line.

centration is also constant, the reaction and electro injection conditions can be maintained with a high percentage of successful reads. Two steps were included to reduce template concentration fluctuation. The first is the introduction of overnight growth of the bacterial clones, resulting in stationary phase cultures irrespective of the amount of bacteria picked for seeding. The second step is to use small amounts of PCR primers. Thirty-five cycles of PCR, using plasmids of high copy number as templates, results in the amount of primers being limiting in later cycles. This eliminates the carryover of amplification primers, avoiding the need for post-PCR purifications. When primer carryover occurs, the template is sequenced simultaneously from both ends, and no useful sequence data is generated. The final effect of our protocol is a homogenization in the quantity of the PCR sequencing templates. The final cost for template generation using this protocol, including all plastics and reagents required, is about US $0.09, a cost significantly lower than previously reported protocols (7,8).

The use of PCR products as templates for capillary sequencing avoids problems of carryover of charged contaminants that may interfere with electroinjection. However, it should be noted that *Taq* DNA polymerase has an estimated error rate of 1/100000 (4), and this should be taken into account. The use of reaction volumes of 15 µL or less, for template generation has an important impact on reducing the final sequencing price by reducing the amounts of reaction consumables. Approximately 5% of the PCRs fail to generate a product, which is probably due to problems in colony picking or colony growth.

This capillary sequencing protocol can be used in conjunction with automated apparatus for colony picking and PCR setup. There is no need for purchasing additional equipment, columns or reagents. The reduction or elimination of a number of cumbersome phases allowed us to achieve a dramatic reduction in costs and an increase in the number of bases sequenced by a full-time employee.

REFERENCES

1. **Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li et al.** 2000. The genome sequence of *Drosophila melanogaster*. Science *287*:2185-2195.
2. **Bonalume Neto, R.** 1999. Brazilian scientists team up for cancer genome project. Nature *398*:450.
3. **Dias Neto, E., C.R. Garcia, S. Verjovski-Almeida, M.R. Briones, M.A. Nagai, W.A. Silva Jr., M.A. Zago, S. Bordin et al.** 2000. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. Proc. Natl. Acad. Sci. USA *97*:3491-3496.
4. **Eckert, K.A. and T.A. Kunkel.** 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res. *18*:3739-3744.
5. **Ewing, B. and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. *8*:186-194.
6. **Ewing, B., L. Hillier, M.C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. *8*:175-185.
7. **Itoh, M., P. Carninci, S. Nagaoka, N. Sasaki, Y. Okazaki, T. Ohsumi, M. Muramatsu, and Y. Hayashizaki.** 1997. Simple and rapid preparation of plasmid template by a filtration method using microtiter filter plates. Nucleic Acids Res. *25*:1315-1316.
8. **Konecki, D.S. and J.J. Phillips.** 1998. Turbo-Prep II: an inexpensive, high-throughput plasmid template preparation protocol. BioTechniques *24*:286-293.
9. **Lee, L.G., C.R. Connell, S.L. Woo, R.D. Cheng, B.F. McArdle, C.W. Fuller, N.D. Halloran, and R.K. Wilson.** 1992. DNA sequencing with dye-labeled terminators and T7 DNA polymerase: effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. Nucleic Acids Res. *20*:2471-2483.
10. **Prober, J.M., G.L. Trainor, R.J. Dam, F.W. Hobbs, C.W. Robertson, R.J. Zagursky, A.J. Cocuzza, M.A. Jensen, and K. Baumeister.** 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides Science *238*:336-341.
11. **Sanger F., S. Nicklen, and A.R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*:5463-5467.
12. **Simpson, A.J. and J.F. Perez.** 1998. ONSA, the Sao Paulo Virtual Genomics Institute. Organization for Nucleotide Sequencing and Analysis. Nat. Biotechnol. *16*:795-796.

We thank Adriana A. Marques, Cristiane Ayres Ferreira, and Alexandre C. de Oliveira for dedicated technical assistance. This work was funded by PRONEX, Fundação de Amparo à Pesquisa do Estado de São Paulo, Ludwig Institute for Cancer Research, and Fundação Hemocentro de Ribeirão Preto. Address correspondence to Dr. Emmanuel Dias Neto, Ludwig Institute for Cancer Research, R. Prof. Antônio Prudente, 109-4th floor, 01509-010, São Paulo, SP Brazil. e-mail: emmanuel@ludwig.org.br

Received 5 July 2000; accepted 13 October 2000.

W.A. Silva, Jr., M.C.R. Costa, V. Valente, J. de Freitas Sousa, M.L. Paçó-Larson, E.M. Espreafico, S.S. Camargo[1], E. Monteiro[1], A. de Jesus Holanda, M.A. Zago, A.J.G. Simpson[1], and E.D. Neto[1]

Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo
[1]Ludwig Institute for Cancer Research
São Paulo, Brazil